NUCLEIC ACID SEQUENCES FROM *DROSOPHILA MELANOGASTER* THAT ENCODE PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF

This application claims the benefit of U.S. Provisional Application No. 60/176,418, filed January 14, 2000, which is hereby incorporated by reference in its entirety.

The Sequence Listing associated with the instant disclosure has been submitted as a 1.37 megabyte file on CD-R (in duplicate) instead of on paper. Each CD-R is marked in indelible ink to identify the Applicants, Title, File Name (31133A.ST25.txt), Creation Date (January 12, 2001), Computer System (IBM-PC/MS-DOS/MS-Windows), and Docket No. (PB/5-31133A). The Sequence Listing submitted on CD-R is hereby incorporated by reference into the instant disclosure.

FIELD OF INVENTION

The present invention pertains to nucleic acid sequences isolated from *Drosophila melanogaster* that encode proteins essential for larval viability. The invention particularly relates to methods of using these proteins as insecticide targets, based on this essentiality.

BACKGROUND OF THE INVENTION

Insects contribute or cause many human and animal diseases, and are responsible for substantial agricultural and property damage. The societal costs associated with insect pests in dollars, time and suffering are monumental. The total worldwide market size for insecticide crop protection is over $5 billion. To combat these problems, insecticidal compounds have been developed and employed.

The idea to use chemicals for insect control is not new. The scientific use of pesticides started with the introduction of arsenical insecticides and organic compounds such as tar, petroleum oils, and dinitrophenol emulsions at the end of the last century. But, the systematic search for synthetic organic insecticides was only launched after the discovery of the insecticidal properties of DDT in 1939. After World War II, chemical research concentrated mainly on

chlorinated hydrocarbons and cyclodienes, which all require high rates of application and have a rather broad spectrum of activity. Most of them are persistent in the environment and may pose a significant risk for accumulation in the food chain. Today the use of these chemicals is very much restricted.

From this point, the major emphasis in research has been given to organophosphates and carbamates, which are readily degradable in the environment with little tendency for bioaccumulation. The toxicity of these compounds varies within a broad range from medium to highly toxic. Organophosphates and carbamates are still widely use, although the more toxic ones are banned in certain countries. The formamidines have as their major advantage a different mode of action and their selectivity, which made them suitable for use in IPM (insect pest management) programs. They are easily degradable with no accumulation potential, but for toxicological reasons some have had to be withdrawn from the market.

For the past decade, insecticide research has concentrated on leadfinding for new chemical structures interfering with new target mechanisms. The chances for success are rather remote, because the hurdles for the registration of a new insecticide are set very high. Toxicological aspects, insecticide resistance, environmental behavior, and IPM fitness are some of the critical factors that have to be considered together with economical factors.

Novel insecticides can now be discovered using high-throughput screens that implement recombinant DNA technology. Proteins found to be essential to insect viability can be recombinantly produced through standard molecular biological techniques and utilized as insecticide targets in screens for novel inhibitors of the enzymes' activity. The novel inhibitors discovered through such screens may then be used as insecticides to control undesirable insect infestation.

However, as the world population continues to grow, there will be increasing food shortages. Therefore, there exists continuing need to find new, effective and economic insecticides.

SUMMARY OF THE INVENTION

In view of these needs, it is one object of the invention to provide essential genes in insects such as *Drosophila melanogaster*. It is another object to provide the essential proteins encoded by these essential genes for assay development to identify inhibitory compounds with insecticidal activity. It is still another object of the present invention to provide an effective and beneficial method for identifying new or improved insecticides using the essential proteins of the invention.

In furtherance of these and other objects, the present invention provides DNA molecules comprising nucleotide sequences isolated from *Drosophila melanogaster* that encode proteins essential for larval viability. The inventors are the first to demonstrate that the nucleotide sequences of the invention are essential for larval viability. This knowledge is exploited to provide novel insecticide modes of action. One advantage of the present invention is that the proteins encoded by the essential nucleotide sequences provide the bases for assays designed to easily and rapidly identify novel insecticides.

Disruption of the nucleotide sequences of the invention demonstrates that the activity of each corresponding encoded protein is essential for *Drosophila* larval viability. Genetic results show that when each nucleotide sequence of the invention is mutated in *Drosophila*, the resulting phenotype is larval lethal in the homozygous state. This demonstrates a critical role for the protein encoded by the mutated nucleotide sequence. This further implies that chemicals that inhibit the expression of the protein when in contact with insects are likely to have detrimental effects on insects and are potentially good insecticide candidates. The present invention therefore provides methods of using the disclosed nucleotide sequences or proteins encoded thereby to identify inhibitors thereof. The inhibitors can then be used as insecticides to kill undesirable insect populations where crops are grown, particularly agronomically important crops such as maize, and other cereal crops such as wheat, oats, rye, sorgum, rice, barley, millet, turf and forage grasses and the like, as well as cotton, sugar cane, sugar beet, oilseed rape, soybeans, vegetable crops and fruits.

The present invention accordingly provides cDNA sequences derived from *Drosophila melanogaster*. In one embodiment, the present invention provides an isolated DNA molecule

comprising a nucleotide sequence selected from the group consisting of the even numbered SEQ ID NOs:14-360. In another embodiment, the present invention provides an isolated DNA molecule comprising a nucleotide sequence that encodes a protein selected from the group consisting of the odd numbered SEQ ID NOs:15-361.

The present invention also provides a chimeric construct comprising a promoter operatively linked to a DNA molecule according to the present invention, wherein the promoter is preferably functional in a eukaryote, wherein the promoter is preferably heterologous to the DNA molecule. The present invention further provides a recombinant vector comprising a chimeric construct according to the present invention, wherein said vector is capable of being stably transformed into a host cell. The present invention still further provides a host cell comprising a DNA molecule according to the present invention, wherein said DNA molecule is preferably expressible in the cell. The host cell is preferably selected from the group consisting of an insect cell, a yeast cell, and a prokaryotic cell.

The present invention also provides proteins essential for *Drosophila melanogaster* larval viability. In one embodiment, the present invention provides an isolated protein comprising an amino acid sequence selected from the group consisting of the odd numbered SEQ ID NOs:15-361. In accordance with another embodiment, the present invention also relates to the recombinant production of proteins of the invention and methods of using the proteins of the invention in assays for identifying compounds that interact with the protein.

In another preferred embodiment, the present invention describes a method for identifying chemicals having the ability to inhibit the activity of the disclosed proteins. In a preferred embodiment, the present invention provides a method for selecting compounds that interact with a protein of the invention, comprising: (a) expressing a DNA molecule according to the present invention to generate the corresponding protein of the invention, (b) testing a compound suspected of having the ability to interact with the protein expressed in step (a), and (c) selecting compounds that interact with the protein in step (b).

Other objects and advantages of the present invention will become apparent to those skilled in the art and from a study of the following description of the invention and non-limiting examples. The entire contents of all publications mentioned herein are hereby incorporated by reference.

BRIEF DESCRIPTION OF THE SEQUENCES IN THE SEQUENCE LISTING

SEQ ID NOs:1-13 are PCR primers.

Even numbered SEQ ID NOs:14-360 are nucleotide sequences described in the table below.

Odd numbered SEQ ID NOs:15-361 are protein sequences encoded by the immediately preceding nucleotide sequence, e.g., SEQ ID NO:15 is the protein encoded by the nucleotide sequence of SEQ ID NO:14, SEQ ID NO:17 is the protein encoded by the nucleotide sequence of SEQ ID NO:16, etc.

SEQ ID NO:	CT	Gene Name	BLAST
14	CT7922	late bloomer (lbl)	H. sapiens 4507541 9.e-05 48.4 gi|4507541|ref|NP_003261.1|| transmembrane 4 superfamily member 6 >gi|2829196 (AF043906) T245 protein
16			
18	CT23784	CG7840	C. elegans 3873678 7.e-35 148 gi|3873678|emb|CAA94885.1| (Z71178) Similarity with yeast hypothetical protein (Swiss prot accession
20			
22	CT23155		
24	CT1431	CG1070	C. elegans 465904 5.e-16 87.8 gi|465904|sp|P34447|YMA2_CAEEL HYPOTHETICAL 56.5 KD PROTEIN F54F2.2 IN CHROMOSOME III >gi|630637|pir
26	CT13858	twinstar (tsr)	C. elegans 4262577 4.e-20 175.6 gi|4262577|gb|AAD14704| (AF125953) contains similarity to C. elegans actin depolymerizing factor UNC
28	CT12433		
30	CT41667	ornithine decarboxylase antizyme	
32	CT41667	GUTFEELING PROTEIN	2.e-12 74.5 gi|1709427|sp|P54368|OAZ_HUMAN ORNITHINE DECARBOXYLASE ANTIZYME (ODC-AZ) >gi|2576244|dbj|BAA23101| (
34	CT24701	small nuclear ribonucleoprotein Sm D3 (from EST)	H. sapiens 4759160 2.e-37 156 gi|4759160|ref|NP_004166.1|| small nuclear ribonucleoprotein D3 polypeptide (18kD) >gi|1173456|sp|P4
36	CT3681	Deadpan Protein	122214 2.e-38 161 gi|122214|sp|P29303|HAIR_DROVI HAIRY PROTEIN >gi|157590 (M87885) basic-helix-loop-helix protein [Drosophila virilis
38		Hsp70/Hsp90 organizing protein homolog	
40	CT13570	DnaJ60	6323870 4.e-08 59.3 gi|6323870|ref|NP_013941.1|SCJ1| dnaJ

			homolog; Scj1p >gi|134297|sp|P25303|SCJ1_YEAST SCJ1 PROTEIN >g
42			
44	CT20524	rab-protein 6	H. sapiens 4506373 2.e-99 363 gi|4506373|ref|NP_002860.1|| RAB6, member RAS oncogene family >gi|131796|sp|P20340|RAB6_HUMAN RAS-RE
46	CT29466	CG10496	
48	CT20712	Klingon	7.e-13 196.7 gi|1017427|emb|CAA62189| (X90569) elastic titin [Homo sapiens]
50	CT21638	protein disulfide isomerase	687235 1.e-144 578.2 gi|687235 (U12440) protein disulfide isomerase [Onchocerca volvulus]
52	CT3604	ferritin subunit	2e-11 120532 P19976 FRI_SOYBN FERRITIN PRECURSOR (SOF-35) >gi|81773|pir||A40992 ferritin precursor -
54		RM62 mRNA for novel RNA helicase	
56	CT3224	string, CDC25 involved in cell cycle control	266557 5.e-52 207 gi|266557|sp|P30309|MPI1_XENLA M-PHASE INDUCER PHOSPHATASE 1/B >gi|419980|pir||A42679 protein-tyrosi
58	CT23521		
60	CT5673	tramtrack p69	H. sapiens 4650844 9.e-08 60.5 gi|4650844|dbj|BAA77027.1| (AB026190) Kelch motif containing protein [Homo sapiens]
62	GM03018		
64	CT2210	UDP-Glucose 4-Epimerase	H. sapiens 4503891 1.e-125 448 gi|4503891|ref|NP_000394.1|| galactose-4-epimerase, UDP- >gi|2494659|sp|Q14376|GALE_HUMAN UDP-GLUCOS
66			
68		ferritin subunit 1 (Fer1)	
70	CT38280	SNF4/AMP-activated protein kinase gamma subunit	
72	CT27760	CG9829	
74	CT13958	heat shock protein cognate 70 (Hsc4)	Remainders 662802 0 1114 gi|662802 (U20256) heat shock-like protein, similar to heat shock 70 kDa proteins [Ceratitis capitat
76		probable transcriptional regulator dre4	
78	CT23870	punt receptor serine/threonine kinase	4406075 1.e-123 444 gi|4406075|gb|AAD19844| (AF069500) activin receptor IIB [Danio rerio]
80		histone 4 replacement gene	
82	CT32834	CG17173	H. sapiens 5748487 4.e-57 222 gi|5748487|dbj|BAA83464.1| (AB000624) UDP-N-acetylglucosamine: alpha-1,3-D-mannoside beta-1,4-N-acet
84	CT7286	abnormal wing disc	2827444 2.e-65 249 gi|2827444 (AF043542) nucleoside diphosphate kinase [Gallus gallus]
86	CT7116	CG2922	H. sapiens 286001 1.e-121 436 gi|286001|dbj|BAA02795| (D13630) KIAA0005 [Homo sapiens]
88	CT16413	head involution defective protein (hid)	D. melanogaster 2498442 1.e-146 521 gi|2498442|sp|Q24106|HID_DROME HEAD INVOLUTION DEFECTIVE PROTEIN (WRINKLED PROTEIN)
90	CT20570	adenylate kinase (ATP-	6707707 2.e-58 226 gi|6707707|sp|Q9WTP7|KAD3_MOUSE

		AMP Transphosphorylase)	GTP:AMP PHOSPHOTRANSFERASE MITOCHONDRIAL (AK3) >gi|4760600|dbj|BAA77
92	CT38193	Drosophila melanogaster prospero gene	H. sapiens 4506119 8.e-53 265.4 gi|4506119|ref|NP_002754.1|| prospero-related homeobox 1 >gi|3024449|sp|Q92786|PRX1_HUMAN HOMEOBOX P
94	CT13750	nonstop , not	
96	CT38193		
98	CT33205	serine/threonine protein kinase	
100			
102		Drosophila melanogaster laminin A chain gene	
104	CT36397	Drosophila melanogaster fatty acid desaturase	5730154 0 732 gi|5730154|emb|CAB52475.1| (AJ245748) fatty acid desaturase [Drosophila simulans]
106	CT9405		4.e-30 130 gi|6691812|emb|CAB65846.1| (AL133506) /prediction=(method:""genscan"", version:""1.0"", score:""109.
108	CT15810	CG4919	H. sapiens 4504011 3.e-19 97.1 gi|4504011|ref|NP_002052.1|| glutamate-cysteine ligase regulatory protein; gamma-glutamylcysteine sy
110			
112		D.melanogaster E2F	
114	CT39616	cytochrome p450 monooxygenase (Cyp6a8)	0.69 AB018267 g3882168 Human mRNA for KIAA0724 protein, complete cds. 0
116	CT1010	Extra-Macrochaetae Protein	
118		frizzled	
120	CT16140	CG7496	4.e-30 131 gi|4827036|ref|NP_005082.1|| TNF superfamily, member 3 (LTB)-like (peptidoglycan recognition protein
122	CT37086	cation-independent mannose-6-phosphate receptor	3.e-41 221.7 gi|3876396|emb|CAB02981.1| (Z81068) similar to LIM domain containing proteins (5 domains); cDNA EST
124	CT37466	pointed ets-like protein (D-ETS-2)	H. sapiens 4885219 8.e-51 316 gi|4885219|ref|NP_005229.1|| v-ets avian erythroblastosis virus E26 oncogene homolog 1 >gi|119641|sp
126	CT22943	osa	
128	CT37466	pointed ets-like prtein (D-ETS-2)	
130	CT42507	D.melanogaster mRNA for serine/threonine protein kinase	8.e-72 455 gi|4505695|ref|NP_002604.1|| 3-phosphoinositide dependent protein kinase-1 >gi|2407613 (AF017995) 3-
132	CT42507	D.melanogaster mRNA for serine/threonine protein kinase	
134	CT36957	Protein Kinase DOA (Protein Darkener Of Apricot	1.e-151 535 gi|4502883|ref|NP_003984.1|| CDC-like kinase 2 isoform hclk2 >gi|1705919|sp|P49760|CLK2_HUMAN PROTEI
136	CT31823	CG11399	(AL137473) hypothetical protein [Homo sapiens], 0.51 J04444 g181239 Human cytochrome c-1 gene

138		Dros Brother protein	
140	CT38229	CG17249	5.e-13 76.5 gi|5174643|ref|NP_005964.1||>gi|2498802|s, 5.1 AF047002 g2896145 Human transcriptional coactivator ALY
142	CT25420	CG8865	H. sapiens 4589562 4.e-94 347 gi|4589562|dbj|BAA76803.1| (AB023176) KIAA0959 protein [Homo sapiens]
144		Drosophila melanogaster mRNA for ferritin subunit 1	
146	CT27454	nuclear distribution gene C	2.e-74 280 gi|1083762|pir||A55897 prolactin-induced T cell protein c15 - rat >gi|619907|emb|CAA57825| (X82445)
148	CT20570	Adenylate Kinase (ATP-AMP Transphosphorylase)	
150	CT18833	ANON-66Da protein	2.e-16 87 gi|1350554|sp|P18615|RDP_HUMAN RD PROTEIN >gi|480387|pir||S36789 gene RD protein - human >gi|190974
152			
154	CT31389	Drosophila melanogaster maelstrom (mael)	
156	CT23870	activin receptor // punt receptor serine/threonine kinase	
158		histone H4	
160	CT40931	Drosophila melanogaster GAGA transcription factor	4.e-07 58.6 gi|3413900|dbj|BAA32314| (AB007938) KIAA0469 protein [Homo sapiens]
162		BcDNA // serine/ threonine protein kinase	
164		H2AvD gene for histone H2A variant	
166	CT20867	CG6719	2.e-50 199 gi|4507873|ref|NP_003363.1|| von Hippel-Lindau binding protein 1 >gi|3212112|emb|CAA76761| (Y17394)
168		IRBP	
170	CT20438	heat shock protein cognate 70	654 gi|6225806|sp|O95757|OS94_HUMAN OSMOTIC STRESS PROTEIN 94 (HEAT SHOCK 70-RELATED PROTEIN APG-1) >gi|
172			
174			
176			
178			
180	CT30041	CG10724	1.e-168 591 gi|3420175|gb|AAD05042| (AF020054) WDR1 protein [Gallus gallus]
182			
184	CT7410	elongation factor 2b	0 1357 gi|3123205|sp|P29691|EF2_CAEEL ELONGATION FACTOR 2 (EF-2) >gi|3876400|emb|CAB02985.1| (Z81068) simil
186	CT14690	argos	1e-122 2708327 (AF038405) argos [Musca domestica]
188	CT8431	Hus-like protein	H. sapiens 4758576 3.e-27 123 gi|4758576|ref|NP_004498.1|| HUS1 (S. pombe) checkpoint homolog >gi|2980665|emb|CAA76518.1| (Y16893)
190			

192	CT21861	pyruvate kinase (Pyk) gene	0 636 gi|125598|sp|P11979|KPY1_FELCA PYRUVATE KINASE, M1 ISOZYME (PYRUVATE KINASE MUSCLE ISOZYME) >gi|8908
194	CT26561	CG9351	1.e-118 428 gi|2315451 (AF016448) No definition line found [Caenorhabditis elegans]
196			
198	CT42625	casein kinase I (dbt)	
200	CT28337	kiwi UDP-glucose dehydrogenase	0 660 gi|4507813|ref|NP_003350.1|| UDP-glucose dehydrogenase >gi|6175086|sp|O60701|UGDH_HUMAN UDP-GLUCOSE
202			
204		G protein alpha subunit gene	
206	CT1010	extramacrochaetae (emc)	
208		(CKI-ALPHA)	
210		ecdysone-inducible gene E75A	
212	CT24290	ecdysone-inducible gene E75B	6166165 0 769.8 gi|6166165|sp|Q08893|E75_MANSE ECDYSONE-INDUCIBLE PROTEIN E75
214		Rga	
216		paramyosin	
218	CT42625	casein kinase I (dbt)	
220			
222			
224	CT23357	phosducin-like protein	4.e-51 202 gi|5912172|emb|CAB56011.1| (AL117602) hypothetical protein [Homo sapiens]
226	CT23245	CG7623	C. elegans 2315363 2.e-59 231 gi|2315363 (AF016441) No definition line found [Caenorhabditis elegans]
228	CT30379	CG10849	6.e-88 324 gi|2144098|pir||I56573 SC2 - rat >gi|256994|bbs|115268 (S45663) SC2=synaptic glycoprotein [rats, bra
230			
232	CT5336	CG12078	H. sapiens 4128029 1.e-38 160 gi|4128029|emb|CAA09865| (AJ011916) hypothetical protein [Homo sapiens]
234		fused protein kinase	
236	CT29244		
238	L08811	daschous (adherin)	
240			
242	CT42625	casein kinase I (dbt)	
244			ABC1 transporter; ABC-type ATPase Magnaporthe grisea
246	CT18629	cyclin A	3.e-72 274 gi|116170|sp|P24861|CG2A_PATVU G2/MITOTIC-SPECIFIC CYCLIN A >gi|84527|pir||S17792 cyclin A - common
248			(DNA (cytosine-5-)-methyltransferase sea urchin)
250	CT22273	fibroblast growth factor receptor homolog DFR1	1.e-113 461.3 gi|558584|emb|CAA68679| (Y00665) tyrosine kinase [Homo sapiens]
252	CT22775	(Ca2+-transporting ATPase chicken)	1026 gi|285369|pir||A42764 Ca2+-transporting ATPase (EC 3.6.1.38) - rat >gi|202862 (M93017) [Rat alternat
254		ald gene for aldolase	
256	CT24935	karyopherin alpha 1	1.e-180 633 gi|4504903|ref|NP_002260.1|| karyopherin alpha

			5; importin alpha 6 >gi|3122273|sp|O15131|IMA6_HUMAN
258		glutamate dehydrogenase	
260		LD21334 unknown mRNA	
262			
264			
266			
268	CT39178	phosphate transporter precursor (MPCP)	1.e-143 508 gi|127276|sp|P16036|MPCP_RAT MITOCHONDRIAL PHOSPHATE CARRIER PROTEIN PRECURSOR (PTP) >gi|112124|pir|
270	CT4906	GAL4 enhancer trap line	6.e-93 340 gi|1154645|emb|CAA64402.1| (X94917) head-elevated expression in 0.9 kb [Drosophila melanogaster]
272		cuticle proteins part of ANT-C gene	
274			
276	CT10611	CG3172	9.e-99 361 gi|6679541|ref|NP_032997.1|| protein tyrosine kinase 9 >gi|1769577 (U82324) A6 protein tyrosine kina
278	CT27808	CG9852, upstream of RpII140	1.e-135 481 gi|85010|pir||JQ1024 hypothetical 30K protein (DmRP140 5' region) - fruit fly (Drosophila melanogast
280	CT18415	CG589, myosin phosphatase like	9.e-86 381.5 gi|633040|dbj|BAA07202| (D37986) 130 kDa myosin-binding subunit of smooth muscle myosin phophatase
282	CT19500	twins, phosphoprotein phosphatase 2A 55 kDa regulatory subunit	0 705.6 gi|4506019|ref|NP_002708.1|| protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), alph
284	CT22097	CG7150	H. sapiens 3258663 4.e-31 217 gi|3258663 (AF064094) KL04P [Homo sapiens]
286	CT24102	CG8045, Su(Raf)3B, d14-3-3epsilon,	Remainders 6634778 6.e-48 342 gi|6634778|gb|AAF19758.1|AC009917_17 (AC009917) Contains similarity to gi|629253 lmbW protein from S
288		STS Dm15553	
290	CT2537	cytochrome B561	4.e-36 153 gi|461668|sp|P34465|C561_CAEEL PUTATIVE CYTOCHROME B561 (CYTOCHROME B-561) >gi|482195|pir||S40988 hy
292	CT23608	CG7769	H. sapiens 1136228 0 1349 gi|1136228 (U32986) UV-damaged DNA binding factor [Homo sapiens] >gi|1588524|prf||2208446A xeroderma
294		catalase (from EST)	
296			
298		Nuclear Pore Complex Protein NUP98	
300	CT21460	LK6 protein kinase (LK6)	1.e-115 419 gi|4464284|gb|AAD21217| (AC007136) Putative map kinase interacting kinase [Homo sapiens] [Homo sapie
302	CT14980	CG4699	Remainders 6331206 5.e-17 91.7 gi|6331206|dbj|BAA86581.1| (AB033093) KIAA1267 protein [Homo sapiens]
304			(nonmuscle myosin heavy chain chicken)
306		Atu	
308	SD10928		
310	CT5324	CG11526	C. elegans 1086830 9.e-91 335 gi|1086830 (U41264) coded for by C. elegans cDNA yk20f8.5; coded for by C. elegans cDNA yk44g1.5; co

312			
314			
316	CT5324	Rel/NF-kappa B homolog (Relish) mRNA	H. sapiens 4502089 5.e-14 271.6 gi|4502089|ref|NP_000028.1|| ankyrin 1, erythrocytic >gi|178646 (M28880) ankyrin [Homo sapiens]
318	CT27692		
320	CT25986	DEAD-box protein	0 881 gi|4826686|ref|NP_004930.1|| DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 1 >gi|6919862|sp|Q92499|DD
322			
324		Dos protein; PH domain protein	
326			
328	CT34516	cytochrome c oxidase subunit Va preprotein	9.e-81 300 gi|6685358|sp|Q94514|COXA_DROME CYTOCHROME C OXIDASE POLYPEPTIDE VA, MITOCHONDRIAL PRECURSOR >gi|165
330	CT22777	effete, UbcD1, CT22799	5.e-83 309 gi|464979|sp|P35129|UBC2_CAEEL UBIQUITIN-CONJUGATING ENZYME E2-17 KD (UBIQUITIN-PROTEIN LIGASE) (UBI
332	CT30099	derailed, CT30136	3.e-72 358.7 gi|1296650|emb|CAA65406| (X96588) receptor tyrosine kinase [Homo sapiens]
334	CT29418	CG10479	
336			
338			
340			
342	CT14836	failed axon connections protein	5.e-28 127 gi|3875233|emb|CAA99802.1| (Z75531) similar to Glutathione S-transferases.; cDNA EST EMBL:T01925 com
344			
346		89B helicase	
348		brahma associated protein 155kDa	
350	CT21153	CG6833	2.e-42 173 gi|1079321|pir||S53818 XPMC2 protein - African clawed frog >gi|595380 (U10185) XPMC2 protein [Xenopu
352			
354	CT22295	CG7228	3.e-75 283 gi|6014727|sp|Q27367|CRQ_DROME CROQUEMORT PROTEIN (D-CD36) >gi|542557|pir||S43137 D-CD36 protein - f
356	CT9784	prophosphoribosylamidotransferase (Prat)	
358			
360			

DEFINITIONS

For clarity, certain terms used in the specification are defined and used as follows:

"Associated with / operatively linked" refer to two nucleic acid sequences that are related physically or functionally. For example, a promoter or regulatory DNA sequence is said to be "associated with" a DNA sequence that codes for an RNA or a protein if the two sequences are

operatively linked, or situated such that the regulator DNA sequence will affect the expression level of the coding or structural DNA sequence.

A "chimeric construct" is a recombinant nucleic acid sequence in which a promoter or regulatory nucleic acid sequence is operatively linked to, or associated with, a nucleic acid sequence that codes for an mRNA or which is expressed as a protein, such that the regulatory nucleic acid sequence is able to regulate transcription or expression of the associated nucleic acid sequence. The regulatory nucleic acid sequence of the chimeric construct is not normally operatively linked to the associated nucleic acid sequence as found in nature.

Co-factor: natural reactant, such as an organic molecule or a metal ion, required in an enzyme-catalyzed reaction. A co-factor is e.g. NAD(P), riboflavin (including FAD and FMN), folate, molybdopterin, thiamin, biotin, lipoic acid, pantothenic acid and coenzyme A, S-adenosylmethionine, pyridoxal phosphate, ubiquinone, menaquinone. Optionally, a co-factor can be regenerated and reused.

A "coding sequence" is a nucleic acid sequence that is transcribed into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Preferably the RNA is then translated in an organism to produce a protein.

Complementary: "complementary" refers to two nucleotide sequences that comprise antiparallel nucleotide sequences capable of pairing with one another upon formation of hydrogen bonds between the complementary base residues in the antiparallel nucleotide sequences.

"Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acid sequences that encode identical or essentially identical amino acid sequences, or where the nucleic acid sequence does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance the codons CGT, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations" which are one species of "conservatively modified variations." Every nucleic acid sequence described herein which encodes a protein also describes every possible silent variation,

except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except ATG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a protein is implicit in each described sequence.

Furthermore, one of skill will recognize that individual substitutions deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations," where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following five groups each contain amino acids that are conservative substitutions for one another: Aliphatic: Glycine (G), Alanine (A), Valine (V), Leucine (L), Isoleucine (I); Aromatic: Phenylalanine (F), Tyrosine (Y), Tryptophan (W); Sulfur-containing: Methionine (M), Cysteine (C); Basic: Arginine (R), Lysine (K), Histidine (H); Acidic: Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q). *See also,* Creighton (1984) *Proteins, W.H.* Freeman and Company. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations."

DNA Shuffling: DNA shuffling is a method to rapidly, easily and efficiently introduce mutations or rearrangements, preferably randomly, in a DNA molecule or to generate exchanges of DNA sequences between two or more DNA molecules, preferably randomly. The DNA molecule resulting from DNA shuffling is a shuffled DNA molecule that is a non-naturally occurring DNA molecule derived from at least one template DNA molecule. The shuffled DNA encodes an enzyme modified with respect to the enzyme encoded by the template DNA, and preferably has an altered biological activity with respect to the enzyme encoded by the template DNA.

Enzyme/Protein Activity: means herein the ability of an enzyme (or protein) to catalyze the conversion of a substrate into a product. A substrate for the enzyme comprises the natural substrate of the enzyme but also comprises analogues of the natural substrate, which can also be converted, by the enzyme into a product or into an analogue of a product. The activity of the enzyme is measured for example by determining the amount of product in the reaction after a

certain period of time, or by determining the amount of substrate remaining in the reaction mixture after a certain period of time. The activity of the enzyme is also measured by determining the amount of an unused co-factor of the reaction remaining in the reaction mixture after a certain period of time or by determining the amount of used co-factor in the reaction mixture after a certain period of time. The activity of the enzyme is also measured by determining the amount of a donor of free energy or energy-rich molecule (e.g. ATP, phosphoenolpyruvate, acetyl phosphate or phosphocreatine) remaining in the reaction mixture after a certain period of time or by determining the amount of a used donor of free energy or energy-rich molecule (e.g. ADP, pyruvate, acetate or creatine) in the reaction mixture after a certain period of time.

Essential: an "essential" *Drosophila melanogaster* nucleotide sequence is a nucleotide sequence encoding a protein such as e.g. a biosynthetic enzyme, receptor, signal transduction protein, structural gene product, or transport protein that is essential to the growth or survival of the insect.

Expression Cassette: "Expression cassette" as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operatively linked to the nucleotide sequence of interest which is operatively linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. Typically, however, the expression cassette is heterologous with respect to the host, i.e., the particular DNA sequence of the expression cassette does not occur naturally in the host cell and must have been introduced into the host cell or an ancestor of the host cell by a transformation event. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter which initiates transcription only when the host cell is exposed to some particular external stimulus. In

the case of a multicellular organism, such as an insect, the promoter can also be specific to a particular tissue or organ or stage of development.

Gene: the term "gene" is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

Heterologous/exogenous: The terms "heterologous" and "exogenous" when used herein to refer to a nucleic acid sequence (e.g. a DNA sequence) or a gene, refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides.

A "homologous" nucleic acid (e.g. DNA) sequence is a nucleic acid (e.g. DNA) sequence naturally associated with a host cell into which it is introduced.

The terms "identical" or percent "identity" in the context of two or more nucleic acid or protein sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection.

Inhibitor: a chemical substance that inactivates the enzymatic activity of an enzyme (or protein) of interest. The term "insecticide" is used herein to define an inhibitor when applied to an insect at any stage of development.

Insecticide: a chemical substance used to kill or inhibit the growth or viability of insects at any stage of development.

Interaction: quality or state of mutual action such that the effectiveness or toxicity of one protein or compound on another protein is inhibitory (antagonists) or enhancing (agonists).

A nucleic acid sequence is "isocoding with" a reference nucleic acid sequence when the nucleic acid sequence encodes a polypeptide having the same amino acid sequence as the polypeptide encoded by the reference nucleic acid sequence.

An "isolated" nucleic acid molecule or an isolated enzyme is a nucleic acid molecule or enzyme that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated nucleic acid molecule or enzyme may exist in a purified form or may exist in a non-native environment such as, for example, a recombinant host cell.

Mature Protein: protein that is normally targeted to a cellular organelle and from which the transit peptide has been removed.

Minimal Promoter: promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation. In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription.

Modified Enzyme Activity: enzyme activity different from that which naturally occurs in an insect (i.e. enzyme activity that occurs naturally in the absence of direct or indirect manipulation of such activity by man), which is tolerant to inhibitors that inhibit the naturally occurring enzyme activity.

Native: refers to a gene that is present in the genome of an untransformed insect cell.

Naturally occurring: the term "naturally occurring" is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a protein or nucleotide sequence present in an organism (including a virus), which can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory, is naturally occurring.

Nucleic acid: the term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence

also implicitly encompasses conservatively modified variants thereof *(e.g.* degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al., Nucleic Acid Res.* 19: 5081 (1991); Ohtsuka *et al., J. Biol. Chem.* 260: 2605-2608 (1985); Rossolini *et al., Mol. Cell. Probes* 8: 91-98 (1994)). The terms "nucleic acid" or "nucleic acid sequence" may also be used interchangeably with gene, cDNA, and mRNA encoded by a gene.

"ORF" means open reading frame.

Purified: the term "purified," when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least about 50% pure, more preferably at least about 85% pure, and most preferably at least about 99% pure.

Two nucleic acids are "recombined" when sequences from each of the two nucleic acids are combined in a progeny nucleic acid. Two sequences are "directly" recombined when both of the nucleic acids are substrates for recombination. Two sequences are "indirectly recombined" when the sequences are recombined using an intermediate such as a cross-over oligonucleotide. For indirect recombination, no more than one of the sequences is an actual substrate for recombination, and in some cases, neither sequence is a substrate for recombination.

"Regulatory elements" refer to sequences involved in controlling the expression of a nucleotide sequence. Regulatory elements comprise a promoter operatively linked to the nucleotide sequence of interest and termination signals. They also typically encompass sequences required for proper translation of the nucleotide sequence.

Significant Increase: an increase in enzymatic activity that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 2-fold or greater of the activity of the wild-type enzyme in the presence of the inhibitor, more preferably an increase by about 5-fold or greater, and most preferably an increase by about 10-fold or greater.

Substantially identical: the phrase "substantially identical," in the context of two nucleic acid or protein sequences, refers to two or more sequences or subsequences that have at least 60%, preferably 80%, more preferably 90, even more preferably 95%, and most preferably at least 99% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In an especially preferred embodiment, the sequences are substantially identical over the entire length of the coding regions. Furthermore, substantially identical nucleic acid or protein sequences perform substantially the same function.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48: 443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally,* Ausubel *et al., infra).*

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al., J. Mol.*

Biol. 215: 403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information on the world wide web at ncbi.nlm.nih.gov/. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix *(see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89: 10915 (1989)).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences *(see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic* Acid Probes part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Typically, under "stringent conditions" a probe will hybridize to its target subsequence, but to no other sequences.

The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.1 5M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see,* Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 45°C for 15 minutes. An example low stringency wash for a duplex

of, *e.g.*, more than 100 nucleotides, is 4-6x SSC at 40°C for 15 minutes. For short probes (*e.g.*, about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

The following are examples of sets of hybridization/wash conditions that may be used to clone homologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention: a reference nucleotide sequence preferably hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

A further indication that two nucleic acid sequences or proteins are substantially identical is that the protein encoded by the first nucleic acid is immunologically cross reactive with, or specifically binds to, the protein encoded by the second nucleic acid. Thus, a protein is typically substantially identical to a second protein, for example, where the two proteins differ only by conservative substitutions.

The phrase "specifically (or selectively) binds to an antibody," or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the

specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised to the protein with the amino acid sequence encoded by any of the nucleic acid sequences of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with other proteins except for polymorphic variants. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual,* Cold Spring Harbor Publications, New York "Harlow and Lane"), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

A "subsequence" refers to a sequence of nucleic acids or amino acids that comprise a part of a longer sequence of nucleic acids or amino acids (e.g., protein) respectively.

"Synthetic" refers to a nucleotide sequence comprising structural characters that are not present in the natural sequence. For example, an artificial sequence that resembles more closely the G+C content and the normal codon distribution of dicot and/or monocot genes is said to be synthetic.

Substrate: a substrate is the molecule that an enzyme naturally recognizes and converts to a product in the biochemical pathway in which the enzyme naturally carries out its function, or is a modified version of the molecule, which is also recognized by the enzyme and is converted by the enzyme to a product in an enzymatic reaction similar to the naturally-occurring reaction.

Target gene: A "target gene" is any gene in an insect cell. For example, a target gene is a gene of known function or is a gene whose function is unknown, but whose total or partial nucleotide sequence is known. Alternatively, the function of a target gene and its nucleotide sequence are both unknown. A target gene is a native gene of the insect cell or is a heterologous gene that had previously been introduced into the insect cell or a parent cell of said insect cell, for example by genetic transformation. A heterologous target gene is stably integrated in the

genome of the insect cell or is present in the insect cell as an extrachromosomal molecule, e.g. as an autonomously replicating extrachromosomal molecule.

Transformation: a process for introducing heterologous DNA into a cell, tissue, or insect. Transformed cells, tissues, or insects are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof.

"Transformed," "transgenic," and "recombinant" refer to a host organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome of the host or the nucleic acid molecule can also be present as an extrachromosomal molecule. Such an extrachromosomal molecule can be auto-replicating. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof. A "non-transformed," "non-transgenic," or "non-recombinant" host refers to a wild-type organism, e.g., a bacterium or plant, which does not contain the heterologous nucleic acid molecule.

Viability: "viability" as used herein refers to a fitness parameter of an insect. Insects are assayed for their homozygous performance of *Drosophila* larval development, indicating which proteins are indispensable to maintain larval life in *Drosophila*.

DETAILED DESCRIPTION OF THE INVENTION

I. Identification Of Essential *Drosophila melanogaster* Nucleotide Sequences Using Transposable Element Insertion Mutagenesis

As shown in the examples below, the identification of novel nucleotide sequences, as well as the essentiality of the nucleotide sequences for normal insect viability, have been demonstrated in *Drosophila* using P-element transposable insertion mutagenesis. Having established the essentiality of the function of the encoded proteins in *Drosophila* and having identified the nucleotide sequences encoding these essential proteins, the inventors thereby provide an important and sought-after tool for new insecticide development.

A lethal phenotype caused by insertion of a P-element indicates that the affected nucleotide sequence codes for an essential protein in the insect. The characterization of the insertion site

using flanking sequence DNA is needed to associate an individual larval lethal line with specific nucleotide sequences. Genomic DNA adjacent to the 5' and/or 3' end of the P-element from the insertion line is generated using inverse PCR.

II. Determining The Complete Coding Sequences Of The Essential *Drosophila* Nucleotide Sequences

The essential *Drosophila* nucleotide sequences are identified by isolating nucleotide sequences flanking the P-element insertion and aligning that sequence with genomic *Drosophila* sequence obtained from the Celera *Drosophila* database. The protein prediction for each genomic region is obtained by use of an exon algorithm program such as GeneMark. All exon algorithm programs currently used for prediction of proteins are susceptible to inaccuracies, including incomplete predictions of coding sequences, missing alternative splice variants, combining of nearby exons of adjacent genes, and mistranslation at intron-exon borders. The prediction of a complete coding sequence can be confirmed by several methods including polymerase chain reaction (PCR) amplification using the 5' and 3' sequence to verify the message, reverse transcription PCR (rtPCR) using an oligonucleotide internal sequence to identify the 5' and/or 3' end, and screening of cDNA libraries from insect tissues with probes made from a particular sequence to isolate a true full-length clone. To confirm that the message size is accurate, a Northern blot can be hybridized with a probe from the nucleotide sequence. In addition, matches to the *Drosophila* EST database helps to confirm existence of message and gives information about the temporal and spatial pattern of expression. Mutation-causing P elements are known to preferentially cluster in the 5' region of affected genes (Spradling *et al., Proc. Natl. Acad. Sci. USA* 92: 10824-10830 (1995)), a tendency that increases the chance of recovering overlaps between short flanking sequences and 5' ESTs. The present invention therefore provides a number of essential nucleotide sequences as well as the amino acid sequences encoded thereby. cDNA clone sequences are set forth in even numbered SEQ ID NOs:14-360. The corresponding encoded amino acid sequences are set forth in odd numbered SEQ ID NOs:15-361.

The isolated gene sequences disclosed herein may be manipulated according to standard genetic engineering techniques to suit any desired purpose. For example, an entire *Drosophila*

gene sequence or portions thereof may be used as a probe capable of specifically hybridizing to coding sequences and messenger RNAs. To achieve specific hybridization under a variety of conditions, such probes include, e.g. sequences that are unique among insect nucleotide sequences for a particular protein of interest and are at least 10 nucleotides in length, preferably at least 20 nucleotides in length, and most preferably at least 50 nucleotides in length. Such probes are used to amplify and analyze related nucleotide sequences from a chosen organism via PCR. This technique is useful to isolate additional insect nucleotide sequences from a desired organism or as a diagnostic assay to determine the presence of particular nucleotide sequences in an organism. This technique also is used to detect the presence of altered nucleotide sequences associated with a particular condition of interest such as insecticide tolerance, poor health, etc.

Gene-specific hybridization probes also are used to quantify levels of a particular gene mRNA in an insect using standard techniques such as Northern blot analysis. This technique is useful as a diagnostic assay to detect altered levels of gene expression that are associated with particular conditions such as enhanced tolerance to insecticides that target a particular gene.

III. Recombinant Production Of Protein And Uses Thereof

For recombinant production of a protein of the invention in a host organism, a nucleotide sequence encoding the protein is inserted into an expression cassette designed for the chosen host and introduced into the host where it is recombinantly produced. The choice of the specific regulatory sequences such as promoter, signal sequence, 5' and 3' untranslated sequence, and enhancer appropriate for the chosen host is within the level of the skill of the routineer in the art. The resultant molecule, containing the individual elements linking in the proper reading frame, is inserted into a vector capable of being transformed into the host cell. Suitable expression vectors and methods for recombinant production of proteins are well known for host organisms such as *E. coli*, yeast, and insect cells (see, e.g., Lucknow and Summers, *Bio/Technol.* 6:47 (1988)). Additional suitable expression vectors are baculovirus expression vectors, e.g., those derived from the genome of *Autographica californica* nuclear polyhedrosis virus (AcMNPV). A preferred baculovirus/insect system is PVL1392(3) used to transfect *Spodoptera frugiperda* SF9 cells (ATCC) in the presence of linear *Autographica californica* baculovirus DNA (Phramingen,

San Diego, CA). The resulting virus is used to infect HighFive *Tricoplusia ni* cells (Invitrogen, La Jolla, CA).

Recombinantly produced proteins are isolated and purified using a variety of standard techniques. The actual techniques used vary depending upon the host organism used, whether the protein is designed for secretion, and other such factors. Such techniques are well known to the skilled artisan (*see, e.g.* chapter 16 of Ausubel, F. *et al.*, "Current Protocols in Molecular Biology", pub. by John Wiley & Sons, Inc. (1994).

IV. Assays For Characterizing The Proteins

Recombinantly produced proteins are useful for a variety of purposes. For example, they can be used in *in vitro* assays to screen known insecticidal chemicals whose target has not been identified to determine if they inhibit protein activity. Such *in vitro* assays may also be used as more general screens to identify chemicals that inhibit such protein activity and that are therefore novel insecticide candidates. Recombinantly produced proteins may also be used to elucidate the complex structure of these molecules and to further characterize their association with known inhibitors in order to rationally design new inhibitory insecticides. Alternatively, the recombinant protein can be used to isolate antibodies or peptides that modulate the activity and are useful in transgenic solutions.

V. *In vivo* Inhibitor Assay: Discovery of Small Molecule Ligands That Interact with Proteins Of Unknown Function.

Having identified a protein as a potential insecticide target based on its essentiality for insect larval viability, a next step is to develop an assay that allows screening large numbers of chemicals to determine which ones interact with the protein. Although it is straightforward to develop assays for proteins of known function, developing assays with proteins of unknown functions can be more difficult.

To address this issue, novel technologies are used that can detect interactions between a protein and a ligand without knowing the biological function of the protein. A short description of three methods is presented, including fluorescence correlation spectroscopy, surface-enhanced laser desorption/ionization, and biacore technologies. In addition to those descibed here, there

are additional methods that are currently being developed that are also amenable to automated, large-scale screening.

Fluorescence Correlation Spectroscopy (FCS) theory was developed in 1972 but it is only in recent years that the technology to perform FCS became available (Madge et al. (1972) *Phys. Rev. Lett.*, 29: 705-708; Maiti et al. (1997) *Proc. Natl. Acad. Sci. USA*, 94: 11753-11757). FCS measures the average diffusion rate of a fluorescent molecule within a small sample volume. The sample size can be as low as 10^3 fluorescent molecules and the sample volume as low as the cytoplasm of a single bacterium. The diffusion rate is a function of the mass of the molecule and decreases as the mass increases. FCS can therefore be applied to protein-ligand interaction analysis by measuring the change in mass and therefore in diffusion rate of a molecule upon binding. In a typical experiment, the target to be analyzed is expressed as a recombinant protein with a sequence tag, such as a poly-histidine sequence, inserted at the N- or C-terminus. The expression takes place in *E. coli*, yeast or insect cells. The protein is purified by chromatography. For example, the poly-histidine tag can be used to bind the expressed protein to a metal chelate column such as Ni2+ chelated on iminodiacetic acid agarose. The protein is then labeled with a fluorescent tag such as carboxytetramethylrhodamine or BODIPY® (Molecular Probes, Eugene, OR). The protein is then exposed in solution to the potential ligand, and its diffusion rate is determined by FCS using instrumentation available from Carl Zeiss, Inc. (Thornwood, NY). Ligand binding is determined by changes in the diffusion rate of the protein.

Surface-Enhanced Laser Desorption/Ionization (SELDI) was invented by Hutchens and Yip during the late 1980's (Hutchens and Yip (1993) *Rapid Commun. Mass Spectrom.* 7: 576-580). When coupled to a time-of-flight mass spectrometer (TOF), SELDI provides means to rapidly analyze molecules retained on a chip. It can be applied to ligand-protein interaction analysis by covalently binding the target protein on the chip and analyze by MS the small molecules that bind to this protein (Worrall et al. (1998) *Anal. Biochem.* 70: 750-756). In a typical experiment, the target to be analyzed is expressed as described for FCS. The purified protein is then used in the assay without further preparation. It is bound to the SELDI chip either by utilizing the poly-histidine tag or by other interaction such as ion exchange or hydrophobic interaction. The chip thus prepared is then exposed to the potential ligand via, for example, a delivery system able to pipet the ligands in a sequential manner (autosampler). The chip is then

submitted to washes of increasing stringency, for example a series of washes with buffer solutions containing an increasing ionic strength. After each wash, the bound material is analyzed by submitting the chip to SELDI-TOF. Ligands that specifically bind the target will be identified by the stringency of the wash needed to elute them.

Biacore relies on changes in the refractive index at the surface layer upon binding of a ligand to a protein immobilized on the layer. In this system, a collection of small ligands is injected sequentially in a 2-5 microlitre cell with the immobilized protein. Binding is detected by surface plasmon resonance (SPR) by recording laser light refracting from the surface. In general, the refractive index change for a given change of mass concentration at the surface layer is practically the same for all proteins and peptides, allowing a single method to be applicable for any protein (Liedberg et al. (1983) *Sensors Actuators* 4: 299-304; Malmquist (1993) *Nature* 361: 186-187). In a typical experiment, the target to be analyzed is expressed as described for FCS. The purified protein is then used in the assay without further preparation. It is bound to the Biacore chip either by utilizing the poly-histidine tag or by other interaction such as ion exchange or hydrophobic interaction. The chip thus prepared is then exposed to the potential ligand via the delivery system incorporated in the instruments sold by Biacore (Uppsala, Sweden) to pipet the ligands in a sequential manner (autosampler). The SPR signal on the chip is recorded and changes in the refractive index indicate an interaction between the immobilized target and the ligand. Analysis of the signal kinetics on rate and off rate allows the discrimination between non-specific and specific interaction.

The compounds that are active in the methods disclosed herein may be used to combat agricultural pests such as aphids, locusts, spider mites, and boll weavils as well as such insect pests which attack stored grains and against immature stages of insects living on plant tissue. The compounds are also useful as a nematodicide for the control of agriculturally important soil nematodes and plant parasites.

VI. Production of peptides

Phage particles displaying diverse peptide libraries permits rapid library construction, affinity selection, amplification and selection of ligands directed against an essential protein (H.B. Lowman, *Annu. Rev. Biophys. Biomol. Struct.* 26, 401-424 (1997)). Structural analysis of

these selectants can provide new information about ligand-target molecule interactions and then in the process also provide a novel molecule that can enable the development of new insecticides based upon these peptides as leads.

The invention will be further described by reference to the following detailed examples. These examples are provided for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

EXAMPLES

Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Sambrook, *et al.*, *Molecular Cloning*, eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989) and by T.J. Silhavy, M.L. Berman, and L.W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and by Ausubel, F.M. *et al.*, *Current Protocols in Molecular Biology*, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987). Well known *Drosophila* molecular genetics techniques can be found, for example, in Robert, D.B., *Drosophila, A Practical Approach* (IRL Press, Washington, DC, 1986).

Example 1: Identification Of Larval Lethal Lines

Essential nucleotide sequences are identified through the isolation of lethal mutants defective in larval development. The genetic scheme for mobilization of P-lacW is as performed in Deak *et. al*, *Genetics* 147: 1697-1722 (1997). Additional larval lethal lines are identified and disclosed in Braun, A., B. Lemaitre, *et al., Genetics* 147: 623-634 (1997); Galloni, M. and B. A. Edgar, *Development* 126: 2365-2375 (1999); Gateff, E., *Int. J. Dev. Biol.* 38(4): 565-590 (1994); Mechler, B. M. J. Biosci., *Bangalore* 19(5): 537-556 (1994); Roch, F., F. Serras, *et al., Mol. Gen. Genet.* 257: 103-112 (1998); Russell, M. A., L. Ostafichuk, *et al., Genome* 41: 7-13 (1998); and in Torok, T., G. Tick, *et al. Genetics* 135: 71-80 (1993). Furthermore, the BDGP gene disruption project of single P-element insertions reveals larval lethal lines mutating 25% of vital *Drosophila* genes Spradling, A. C., D. Stern, *et al., Genetics* 153: 135-177 (1999).

Males carrying the transposase source P(Δ2-3) are crossed en masse to yellow white females homozygous for a P-lacW insertion on the X chromosome. Males carrying the PlacW insertion on the X and Δ2-3 on the third chromosome are collected from this cross. The F0 "jumpstart" males are crossed in groups of 10-15 to 20-25 females of w spl; Sb/TM3, Ser genetype. Male F1 progeny with pigmented eyes indicate that the P-lacW has jumped to an autosome. An average of 10-15 males from each F0 cross lacking Δ2-3 are crossed individually to y w; DTS-4/TM3, Sb Ser females, that all third chromosomal insertions result in balanced F2 stocks. Insertions on other autosomes yield white-eyed flies in the F2 generation and are eliminated. The balanced third chromosome insertions are tested for lethality in the next generation by placing four to six pairs of y w; P-lacW/TM3, Sb Ser flies in a vial and examining their progeny for the presence of homozygous P-lacW flies. To analyze the lethal phase, the TM3, Sb Ser balancer is replaced by the TM6C, TB Sb chromosome. In such a genetic background, homozygous mutants can be identified by their wild-type body-length. An average of 10-15 pairs of flies are placed in vials supplemented with yeast paste, and the eggs are collected from each line for 1 day. The development of 50-100 progeny is monitored, and the presence of homozygotes are recorded in all developmental stages. Lethal phase is assigned to a developmental stage in which homozygote animals last appear. Larval lethal lines are identified and maintained.

Example 2: Sequence Determination

Inverse PCR: To determine the flanking sequence of the larval lethal lines, the "Inverse PCR and Cycle Sequencing Protocol for Recovery of Sequences Flanking PZ, PlacW, and PEP elements" of E. Jay Rehm, Berkeley *Drosophila* Genome Project on the world wide web at fruitfly.org/methods/ is used with slight modifications. These modifications include the following: genomic DNA is obtained from 10 flies, rather than 30 flies, with adjustments for final concentrations; all DNA precipitations are performed using glycogen; for some reactions, all of the digest volume is used in the appropriate ligations; the number of cycles in PCR reactions was increased to 40; Pry1 and Pry2 were used to sequence the PEP line flanking sequences.

Genomic DNA isolation: Flies are collected and frozen at –20°C until ready for use. Genomic DNA is prepared by grinding flies in 200 µl Buffer A with a disposable grinder 30X (Buffer A is composed of 100 mM Tris-Cl, pH7.5, 100 mM EDTA, 100 mM NaCl, 0.5% SDS). Add 200 µl additional Buffer A; grind another 15X. Keep on ice until finished. Incubate at 65°C for 30 minutes. Vortex to mix. Add 800 µl freshly made LiCl/KAc Solution (LiCl/ Kac Solution is comprised of 1 part 5 M KAc and 2.5 parts 6 M LiCl). Vortex. Incubate –20°C for 20 minutes. Spin at maximum speed at room temperature 15+ minutes. Transfer 1 ml supernatant to a clean tube avoiding floating debris. Add 600 µl room temperature isopropanol to supernatant. Mix well by tipping. Add 0.5 µl glycogen. Vortex. Incubate at room temperature for 5 minutes. Spin 15 minutes at room temperature, maximum speed. Aspirate away the supernatant. Wash 2X with 500 µl 70% room temperature ethanol; vortex between washes. Spin for 10 minutes at room temperature, maximum speed. Aspirate away supernatant. Dry in a speed vacuum for 10 minutes. Resuspend in 50 µl TE + 0.1 mg/ml RNAse A {for 1 ml TE/RNAse A Solution, add 990 µl TE + 10 µl RNAse A (10mg/ml)). Check 5 µl on 0.8% gel.

Digest Genomic DNA (Sau3A I, HinP1 I, or Msp I--done separately): Set up digests in 96 well tray. Per reaction, add 10 µl genomic DNA, 5 µl 10X Buffer, 2 µl 0.1mg/ml RNAase A stock, 30.5 µl dH_2O, 10 units of enzyme (8 units for Sau 3A I), 0.5µl of 100X BSA (for Sau 3AI only). Incubate at 37°C for 2.5 hours. Check on 0.8% gel before heat-inactivating at 65°C for 20 minutes.

Ligate P Element and Flanking DNA: Set-up ligation tube with 400 µl of ligation mixture then add 30-50 µl of the digest: Per reaction, add 30 µl of digested genomic DNA, 43 µl of 10X ligation buffer (NEB), 375 µl of dH_2O, and 2 µl of ligase (2 Weiss units). Incubate overnight at 4°C. Total reaction volume is adjusted as appropriate.

Precipitate Ligated DNA: To ligation tube, add 40 µl 3M NaAc pH5.2 + 1ml 100% room temperature ethanol + 1 µl glycogen. Mix by tipping. Incubate –20°C for 15+ minutes. Spin 15 minutes, 4°C. Aspirate away supernatant. Wash with 500 µl room temperature 70% ethanol. Vortex. Spin room at temperature for 10 minutes. Aspirate away supernatant. Dry in speed vacuum for 10 minutes. Resuspend in 50 µl TE. Vortex to mix. Transfer to 96 well plate.

PCR: Set up PCR reactions in 96 well plates (Applied Biosystems). Set up PCR reactions with primers appropriate for the type of P element and the end of the element from which

genomic sequence is to be recovered.

Primers for PCR: (type of P element 5' or 3' end forward primer reverse primer annealing temperature):

PZ P-element5' endPlac4Plac1	60°
PZ P-element3' endPry4Pry1	55°
PZ P-element3' endPry2Pry1	60°
PlacW P-element5' endPlac4Plac1	60°
PlacW P-element3' endPry4Plw3-1	55°
PlacW P-element3' endPry2Pry1	60°
PEP P-element5' endPwht1Plac1	60°
PEP P-element3' endPry4Pry1	55°
PEP P-element3' endPry2Pry1	60°

The Pry2/Pry1 combination has a higher annealing temperature than the Pry4/Pry1 and Pry4/Plw3-1 combinations, but the resulting PCR products do not allow sequencing directly off the 3' end of the P-element. The latter primer combinations are therefore used in all initial experiments; the Pry2/Pry1 combination can be used in those cases where strong and unique bands do not result.

Per reaction: 10 µl of ligated genomic DNA, 1 µl of 10mM dNTP mix, 1 µl of 10 µM forward primer stock, 1µl of 10 µM reverse primer stock, 5 µl of 10X Qiagen Taq buffer, 31.5 µl of dH_2O, 0.5 µl of Qiagen Taq.

Cycles: 1X 95°C for 5 minutes; 40X (95°C for 30 seconds; 60°C (high temp) or 55°C (low temp) for 30 seconds; 68°C for 2 minutes); 1X 72°C for 10 minutes; hold at 4°C; run 10µl on 1.5% gel to check. Rearray positive wells to 96 well plate for sequencing clean-up. The primer sets for PCR are as shown in the table below:

Digest, End, Temperature	Forward PCR Primer	Reverse PCR Primer
H5h	Plac4	Plac1
H3h	Pry2	Pry1
H3l	Pry4	Plw3-1
M5h	Plac4	Plac1
M3h	Pry2	Pry1
M3l	Pry4	Plw3-1
S5h	Plac4	Plac1
S3h	Pry2	Pry1
S3l	Pry4	Plw3-1

PCR Primer Sequences (5' to 3'):

Plac4 (27)	- act gtg cgt tag gtc ctg ttc att gtt	SEQ ID NO:1
Plac1 (24)	- cac cca agg ctc tgc tcc cac aat	SEQ ID NO:2
Pry4 (23)	- caa tca tat cgc tgt ctc act ca	SEQ ID NO:3
Pry1 (26)	- cct tag cat gtc cgt ggg gtt tga at	SEQ ID NO:4
Pry2 (28)	- ctt gcc gac ggg acc acc tta tgt tat t	SEQ ID NO:5
Plw3-1 (19)	- tgt cgg cgt cat caa ctc c	SEQ ID NO:6
Pwht1 (19)	- gta acg cta atc act ccg aac agg tca ca	SEQ ID NO:7

Enzymatic Clean-Up for Sequencing: To 40 µl PCR reaction, add 4 µl of enzyme mix. Incubate at 37°C for 1 hour. Inactivate at 70°C for 10 minutes. (Enzyme Mix consists of 2.5U/µl Exonuclease I (Amersham E700732), 0.5U/µl Shrimp Alkaline Phosphatase (Amersham E70183), 1X Amplitaq PCR buffer, add dH_20 to final volume.)

Example 3: Sequence Analysis

Sequence of the flanking sequence generated by inverse PCR is performed on an ABI 3700 sequencer (Perkin Elmer) using BIG DYE sequencing reaction.

Primer sets for sequencing are as shown in the table below:

Digest, End, Temperature	Forward Primer	Reverse Primer
H5h	Splac2	Sp1
H3h	Pry2	Sp5
H3l	Spep1	Sp5
M5h	Splac2	Sp1
M3h	Pry2	Sp5
M3l	Spep1	Sp5
S5h	Splac2	Sp1
S3h	Pry2	Sp6
S3l	Spep1	Sp6

The following primer sets are designed to sequence both ends of PCR products recovered from PlacW and PZ strains:

Splac2 and Sp1 - for use with the Plac4/Plac1 5' PCR primer combination with either PZ or PlacW P-elements; allows sequencing of both ends of the PCR fragment.

Spep1 and Sp3 - for use with the Pry4/Pry1 3' PCR primer combination with PZ P-elements; allows sequencing of both ends of the PCR fragment.

Spep1 and Sp6 - for use with the Pry4/Plw3-1 3' PCR primer combination with PlacW P-elements where Sau3a digestion is performed; allows sequencing of both ends of the PCR fragment.

Spep1 and Sp5 - for use with the Pry4/Plw3-1 3' PCR primer combination where HinP1 digestion is performed; allows sequencing of both ends of the PCR fragment.

Pry1 and Pry2 – for use with the Pry1/Pry2 3' PCR primer combination; allows sequencing of both ends of the PCR fragment.

The PCR products recovered from PEP strains are sequenced with the following primers: Sp1- for use with the Pwht1/Plac1 5' PCR primer combination with the PEP element; Spep1- for use with the Pry4/Pry1 3' PCR primer combination with the PEP element; Pry1 and Pry2 for use with the Pry1/Pry2 3' PCR primer combination with the PEP element.

Primer Sequences (5' to 3'):

Splac2 (25)	- gaa ttc act ggc cgt cgt ttt aca a	SEQ ID NO:8
Sp1 (22)	- aca caa cct ttc ctc tca aca a	SEQ ID NO:9
Sp3 (24)	- gag tac gca aag ctt taa cta tgt	SEQ ID NO:10
Sp6 (23)	- tga cca cat cca aac atc ctc tt	SEQ ID NO:11
Sp5 (25)	- gca tca caa aaa tcg acg ctc aag t	SEQ ID NO:12
Spep1 (19)	- gac act cag aat act att c	SEQ ID NO:13

Melting temperatures of sequencing primers:

Splac2- 60.1°C

Sp1- 50.6°C

Sp3- 49.3°C

Sp6- 54.9°C

Sp5 -60.3°C

Spep1- 44.8°C

Example 4: Secondary Confirmation of Lethality

The lethality of the chromosome carrying the P-element insertion is demonstrated genetically as described in Example 1. The essential *Drosophila* nucleotide sequences are identified by isolating nucleotide sequences flanking the P-element insertion and aligning those sequences with genomic *Drosophila* sequence obtained from the Celera *Drosophila* database. However, in some instances, a second site mutation exists on the chromosome that is responsible for the lethality. In other instances, the location of the flanking sequence is such that determination of which gene(s) are affected by the P-element insertion is rendered difficult or impossible. Thus, to provide secondary confirmation that the gene indicated is essential, there are many methods that one skilled in the art can use, e.g., rescue of the lethality using

transformation technology, perturbation of the gene in a targeted manner, or failure to complement a deficiency.

To provide secondary confirmation, larval lethal lines are crossed to a line containing a deficiency spanning the region of the insert. This creates a hemizygous condition in that particular region and reveals the recessive phenotype of the P-element. Complementation with deficiencies that unequivocally remove the P-element insertion site is taken as proof that the P-element does not cause the associated phenotype. Failure to complement indicates that the strain is verified. This method is as performed in Spradling, A. C., D. Stern, *et al., Genetics* 153: 135-177 (1999). While lines with secondary mutations closely linked to the P insertion might be erroneously verified by this procedure, further molecular and genetic analyses suggest that the frequency of such errors is small. RNA interference, described in Fire, A., S. Xu, *et al., Nature* 391, 806-811 (1998) and Kennerdell, J.R. and Carthew, R.W., *Cell* 95, 1017-1026 (1998), is used as a method to target a gene of interest and demonstrate that the perturbation of the identified gene produces a lethal phenotype.

Example 5: Isolation Of Full Length cDNA

A cDNA screen is performed using a *Drosophila melanogaster* cDNA library probed with a portion of each nucleotide sequence disclosed in the Sequence Listing. Positive colonies are selected, a subset sequenced, and a clone corresponding to the full-length cDNA is recovered. Alternatively, primers from the predicted 5' and 3' end are used in polymerase chain reaction with either a *Drosophila* cDNA library or first strand cDNAs obtained by reverse transcription of *Drosophila* mRNAs as template to amplify a fragment representing the full-length clone.

Example 6: Expression Of Recombinant Protein In Insect Cells

Baculovirus vectors, which are derived from the genome of AcNPV virus, are designed to provide high levels of expression of cDNA in the SF9 line of insect cells (ATCC CRL# 1711). Recombinant baculovirus expressing the cDNA of the present invention is produced by the following standard methods (InVitrogen MaxBac Manual): cDNA constructs are ligated into the

polyhedrin gene in a variety of baclovirus transfer vectors, including the pAC360 and the BleBAc vector (InVitrogen). Recombinant baculoviruses are generated by homologous recombination following co-transfection of the baculovirus transfer vector and linearized AcNPV genomic DNA (Kitts, P.A., *Nucleic Acid. Res.* 18: 5667 (1990)) into SF9 cells. Recombinant pAC360 viruses are identified by the absence of inclusion bodies in infected cells and recombinant pBlueBac viruses are identified on the basis of B-galactosidase expression (Summers, M.D. and Smith, G.E., Texas Agriculture Exp. Station Bulletin No. 1555). Following plaque purification, the *Drosophila* cDNA expression is measured.

The cDNA encoding the entire open reading frame for the *Drosophila* cDNA is inserted into the BamHI site of pBlueBacII. Constucts in the positive orientation, which are identified by sequence analysis, are used to transfect SF9 cells in the presence of linear AcNPV wild type DNA. Authentic, active *Drosophila* cDNA is found in the cytoplasm of infected cells. Active *Drosophila* cDNA is extracted from infected cells by hypotonic or detergent lysis.

Example 7: Expression Of Recombinant Protein In *E. coli*

A cDNA clone of the present invention is subcloned into an appropriate expression vector and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the recombinant protein is confirmed. Recombinant protein is then isolated using standard techniques.

Example 8: *In vitro* Binding Assays

Recombinant protein is obtained, for example according to Example 6 or Example 7. The protein is immobilized on chips appropriate for ligand binding assays. The protein immobilized on the chip is exposed to sample compound in solution according to methods well know in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements

are SEDLI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

The above disclosed embodiments are illustrative. This disclosure of the invention will place one skilled in the art in possession of many variations of the invention. All such obvious and foreseeable variations are intended to be encompassed by the appended claims.